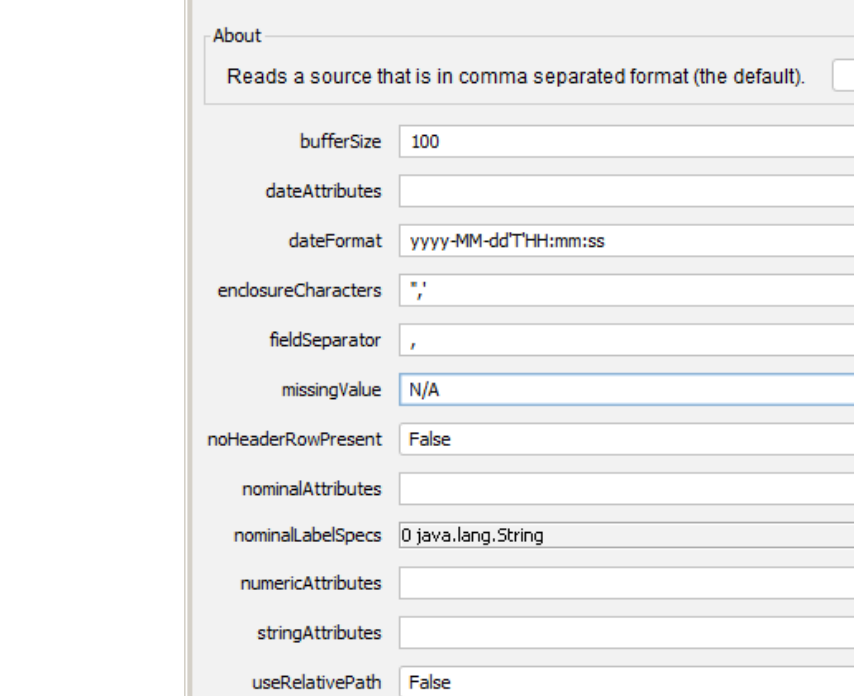


2. Del dataset anterior realice en WEKA, tres algoritmos de preprocesamiento.

[illegible]

The screenshot shows the 'weka.gui.GenericObjectEditor' window for the 'weka.core.converters.CSVLoader' class. The 'About' tab is active, displaying the description: 'Reads a source that is in comma separated format (the default).' with a 'More' button. Below this, various configuration options are listed with their current values:

- bufferSize: 100
- dateAttributes: (empty)
- dateFormat: yyyy-MM-dd'THH:mm:ss
- enclosureCharacters: ",'
- fieldSeparator: ,
- missingValue: N/A
- noHeaderRowPresent: False
- nominalAttributes: (empty)
- nominalLabelSpecs: 0 java.lang.String
- numericAttributes: (empty)
- stringAttributes: (empty)
- useRelativePath: False

At the bottom, there are four buttons: 'Open...', 'Save...', 'OK', and 'Cancel'.

De esta forma los valores nulos se presentaran como espacios vacíos, tal es el ejemplo de la columna Year.

Nombre: Landaeta Navarro Adolfo Antonio

Viewer

Relation: data4

No.	1: i>gRank Numeric	2: Name Nominal	3: Platform Nominal	4: Year Numeric	5: Genre Nominal	6: Publisher Nominal	7: NA_Sales Numeric	8: EU_Sales Numeric	9: JP_Sales Numeric	10: Other_Sales Numeric	11: Global_Sales Numeric
1...	15882.0	Cabela...	PS2		Sports	Unknown	0.01	0.01	0.0	0.0	0.02
1...	15609.0	Writin...	DS		Misc	Unknown	0.0	0.0	0.02	0.0	0.02
1...	15655.0	Virtua ...	PS2		Role-Pl...	Unknown	0.01	0.01	0.0	0.0	0.02
1...	15700.0	Shone...	DS		Misc	Konami Di...	0.0	0.0	0.02	0.0	0.02
1...	15720.0	Witho...	XB		Shooter	Capcom	0.01	0.0	0.0	0.0	0.02
1...	15742.0	Footb...	X360		Sports	Sega	0.0	0.01	0.0	0.0	0.02
1...	15814.0	Ferrari...	Wii		Racing	System 3 ...	0.0	0.01	0.0	0.0	0.02
1...	15868.0	PDC W...	DS		Sports	Unknown	0.01	0.0	0.0	0.0	0.02
1...	15879.0	Dinoto...	GC		Action	Unknown	0.01	0.0	0.0	0.0	0.02
1...	15881.0	Dance...	Wii		Misc	DTP Enter...	0.0	0.01	0.0	0.0	0.02
1...	15903.0	Jet Im...	DS		Simulat...	Nintendo	0.0	0.0	0.02	0.0	0.02
1...	15918.0	Dream...	DS		Misc		0.01	0.0	0.0	0.0	0.02
1...	16060.0	PDC W...	PSP		Sports	Oxygen I...	0.01	0.0	0.0	0.0	0.01
1...	16061.0	WRC: ...	PC		Racing	Black Bea...	0.0	0.01	0.0	0.0	0.01
1...	16068.0	Aquam...	XB		Action	Unknown	0.01	0.0	0.0	0.0	0.01
1...	16194.0	Home...	PC		Strategy		0.0	0.01	0.0	0.0	0.01
1...	16197.0	Shorts	DS		Platform	Unknown	0.01	0.0	0.0	0.0	0.01
1...	16201.0	AKB1/...	X360		Misc		0.0	0.0	0.01	0.0	0.01
1...	16232.0	Brothe...	X360		Shooter		0.01	0.0	0.0	0.0	0.01
1...	16249.0	Agare...	PS3		Role-Pl...	Idea Fact...	0.0	0.0	0.01	0.0	0.01
1...	16310.0	Freaky...	GC		Racing	Unknown	0.01	0.0	0.0	0.0	0.01
1...	16330.0	Inversion	PC		Shooter	Namco Ba...	0.01	0.0	0.0	0.0	0.01
1...	16369.0	Hakuo...	PS3		Advent...	Unknown	0.01	0.0	0.0	0.0	0.01
1...	16430.0	Virtua ...	GC		Role-Pl...	Unknown	0.01	0.0	0.0	0.0	0.01
1...	16496.0	The S...	3DS		Action	Unknown	0.0	0.01	0.0	0.0	0.01

Add instance Undo OK Cancel

Ahora analizando los atributos y que presentan las columnas podemos interpretar que el valor de la columna Year esta clasificada como Numerico, pero debido a quel atributo año es brindar información categorica deberemos cambiarlo, y como el valor de La columna Publisher fue designado como "N/A" y "Unknown" estadarizamos manualmente a que el valor por defecto se "Unknown"

1.

El Primer algoritmo de preprocesamiento “NumericToNominal” que nos ayuda a cambiar el atributo que tenga una columna numérica de Year

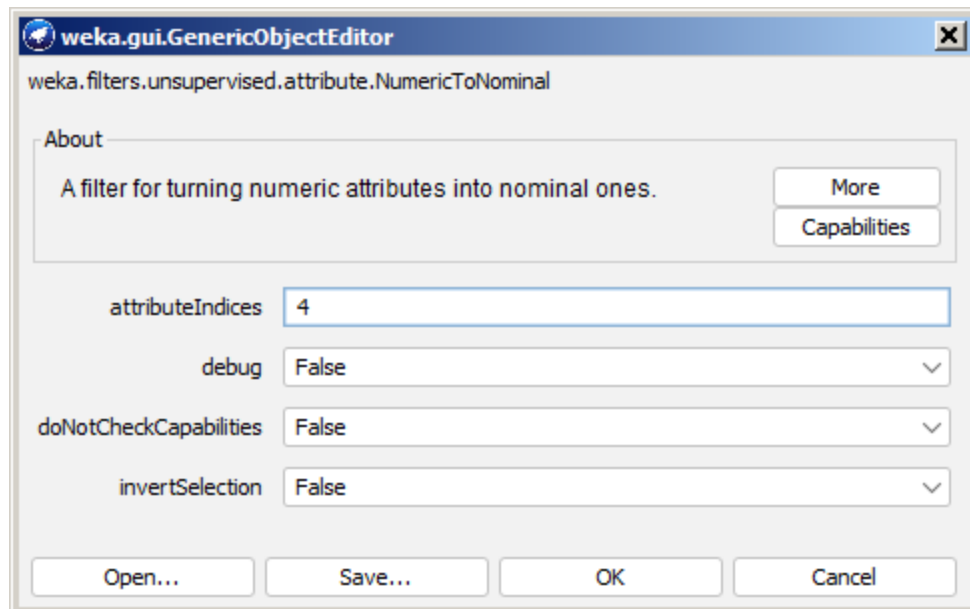
The screenshot shows the Weka Explorer application. The 'Preprocess' tab is active. In the 'Filter' list on the left, 'NumericDiscretize' is selected. The main window displays the 'Global_Sales (Num)' attribute with the following summary:

Selected attribute	
Name: i\oRank	Type: Numeric
Missing: 0 (0%)	Distinct: 16598
Unique: 16598 (100%)	
Statistic	Value
Minimum	1
Maximum	16600
Mean	8300.605
StdDev	4791.854

Below the summary, there is a visualization area showing a bar chart with many bars, all labeled '004'. The status bar at the bottom indicates 'Status OK'.

En su configuración solo tomamos la columna 4 que es la que necesitamos

Nombre: Landaeta Navarro Adolfo Antonio



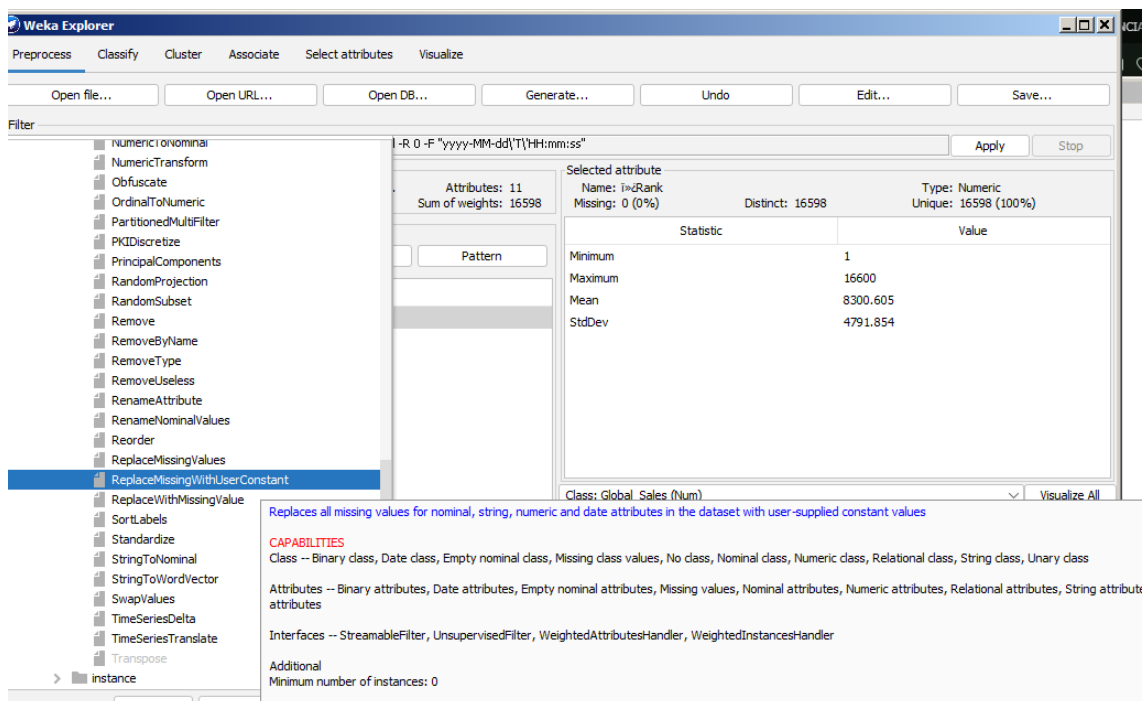
Lo aplicamos y vemos el cambio

The first screenshot shows the 'Viewer' window with 'Relation: data4'. It displays a table with 11 columns: No., 1: i>Rank, 2: Name, 3: Platform, 4: Year, 5: Genre, 6: Publisher, 7: NA_Sales, 8: EU_Sales, 9: JP_Sales, 10: Other_Sales, and 11: Global_Sales. The data is partially visible, showing rows 1 and 2.

The second screenshot shows the 'Viewer' window with 'Relation: data4-weka.filters.unsupervised.attribute.NumericToNominal-R4'. It displays the same table, but the 'Year' attribute (column 4) is now nominal, with values '2006', '1985', and '2008'.

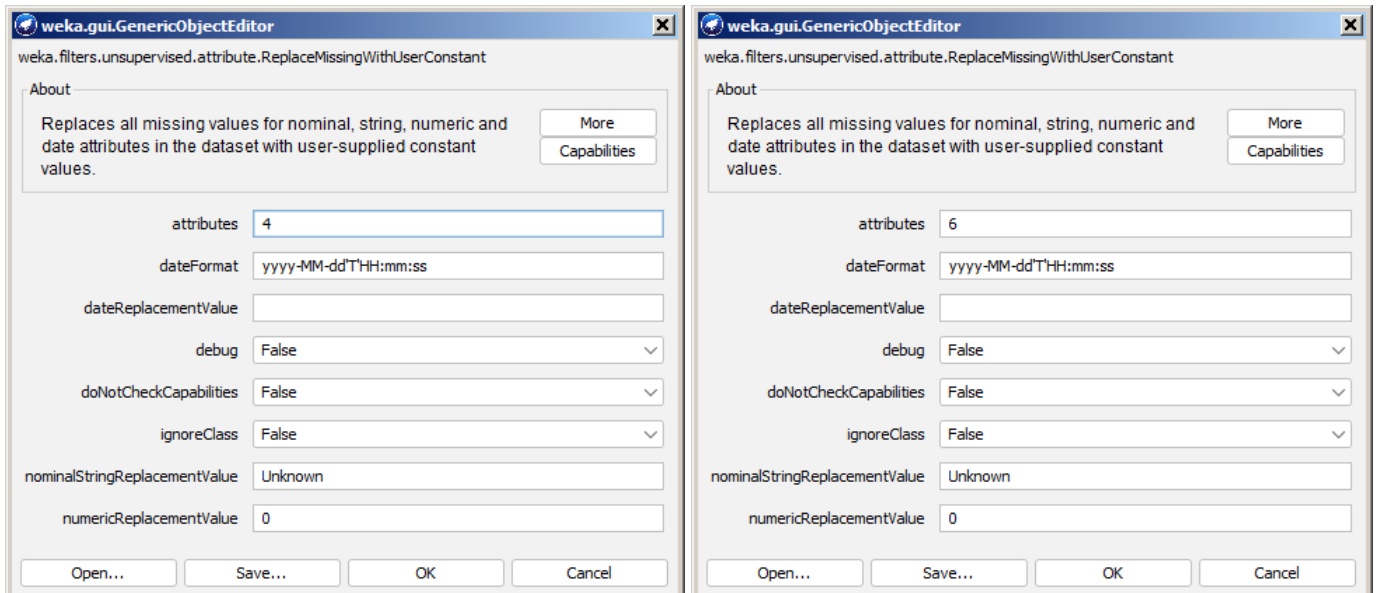
2.

Ya asignado los atributos correctos, usaremos el segundo preprocesamiento "ReplaceMissingWithUserConstant"



Con el cual en los valores perdidos los rellenaremos con el valor estándar Unknown

Nombre: Landaeta Navarro Adolfo Antonio



Choose **ReplaceMissingWithUserConstant** -A 6 -N Unknown -R 0 -F "yyyy-MM-dd\T\HH:mm:ss"

Current relation
Relation: data4-weka.filters.unsupervised.attribute.NumericToNom... Attributes: 11
Instances: 16598 Sum of weights: 16598

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> iRank
2	<input type="checkbox"/> Name
3	<input type="checkbox"/> Platform
4	<input checked="" type="checkbox"/> Year
5	<input type="checkbox"/> Genre
6	<input type="checkbox"/> Publisher
7	<input type="checkbox"/> NA_Sales
8	<input type="checkbox"/> EU_Sales
9	<input type="checkbox"/> JP_Sales

Selected attribute
Name: Year
Missing: 0 (0%)
Distinct: 40
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	Unknown	271	271
2	1980	9	9
3	1981	46	46
4	1982	36	36
5	1983	17	17
6	1984	14	14
7	1985	14	14
8	1986	21	21
9	1987	16	16
10	1988	15	15
11	1989	17	17

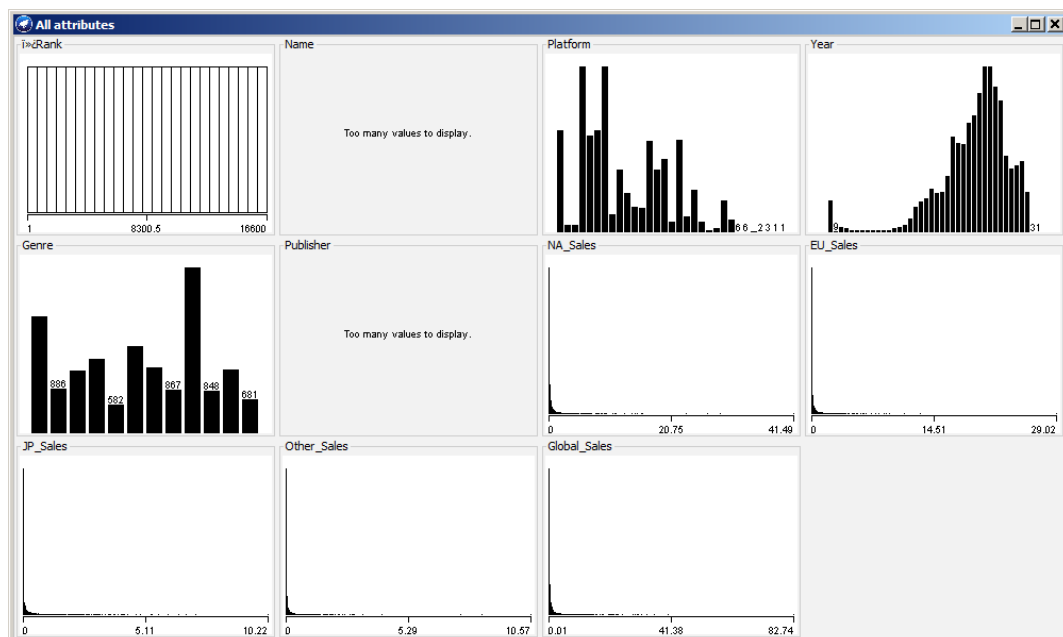
Class: Global_Sales (Num)

Visualize All

Con esta remplazaremos los valores nulos de Year y Publisher.

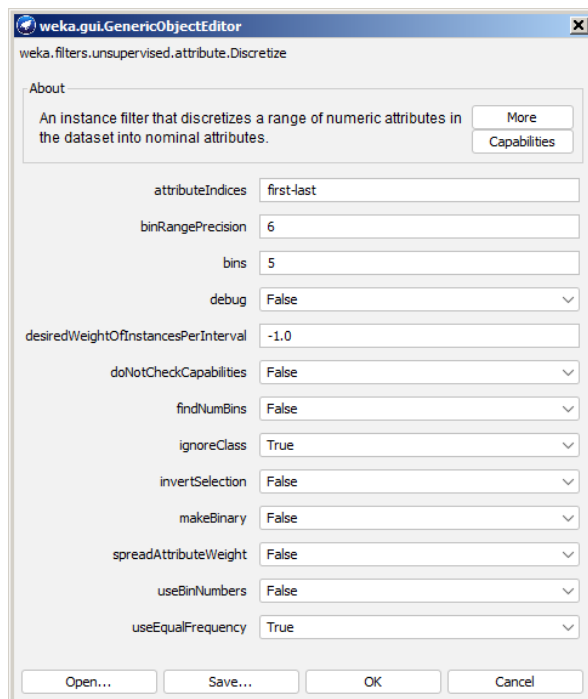
3.

Analizando un poco la visualización de los valores numericos podemos ver que los valores se encuentran demasiado dispersos por lo cual podemos normalizar los valores con "Discretize"



Nombre: Landaeta Navarro Adolfo Antonio

Configuraremos el valor a 5 rangos, manejaremos el uso de frecuencias iguales y ignoremos la clase puesto que con el dataset podemos asignar diferentes columnas como columna de clase



Volviendo a visualizar veremos que los valores se encuentran mejor ordenado

